
Lin AN

Contact: 202-257-5883

Email: lua137@psu.edu

Add: 234 Bradley Ave, State College PA 16801

SUMMARY

- 5 + years of experience in building predictive and explanatory models
- deep understanding of machine learning methods
- advanced coding expertise in Python, R, Shell script
- hands on experience with various NGS data including ChIP-seq, Hi-C, RNA-seq, ATAC-seq and ChIA-PET
- experienced in using public data including NCBI, ENCODE, 4DN and UCSC
- developed multiple downstream analysis tools/pipelines for genomics research

Education Background

Pennsylvania State University	Ph.D. in Bioinformatics – GPA 3.80/4.0	August 2014 - Now
Georgetown University	M.S. in Bioinformatics – GPA 3.95/4.0	August 2013
Northwest Agriculture and Forestry University	B.S. in Bioengineering – GPA 3.59/4.0	June 2012

Experiences

05/2018-08/2018 Computational Biology Intern

Ancestry.com Inc. – San Francisco, CA

- Developed an efficient machine learning pipeline to discover haplotype for genetic communities in AncestryDNA network.

08/2014-Present Ph.D. Candidate

Pennsylvania State University – State College, PA

Advisor: Dr. Yu Zhang, Dr. Feng Yue and Dr. Ross Hardison

OnTAD: An optimal caller for nested TAD

- Developed a computational method that can finish TAD calling in 10mins, which is 5X faster than other published methods.

HiCPlus: Enhancing Hi-C data resolution with deep convolutional neural network

- Co-developed first method to computationally improve the quality of chromatin conformation data, which significantly reduced the cost of sequencing experiments.

Jointly characterizing epigenetic dynamics across multiple cell types

- Assisted with designing and evaluation computational method for multiple cell type chromatin states annotation and provided comprehensive segmentation across 20 hematopoiesis cell types.

06/2013-04/2014 Bioinformatics Analyst

Georgetown Lombardi Comprehensive Cancer Center – Washington, DC

SNP2Structure: A public resource for mapping and modeling nsSNPs on human protein structures

- Integrated 7,021 protein structures with 26,097 related DNA mutations and implemented online portal to visualize effects of SNPs on corresponding protein structures.

07/2013-08/2013 Researcher

Virginia Bioinformatics Institute – Blacksburg, VA

CAN-zip: centroid based compression of next generation sequencing (NGS) data using arithmetic encoding

- Co-designed and implemented a compression software that specialized for NGS data, which reduced file size to 1/10.

Publications

- Yan Zhang*, **Lin An***, Jie Xu, Bo Zhang, W. Jim Zheng, Ming Hu, Jijun Tang# and Feng Yue#. "Enhancing Hi-C data resolution with deep convolutional neural network HiCPlus." *Nature Communications* 2018, (co-first author)
- Jesse Dixon*, Jie Xu*, Vishnu Dileep*, Ye Zhan*, Fan Song, Victoria T. Le, ..., **Lin An**, Ferhat Ay, William Stafford Noble, Job Dekker, David M. Gilbert, Feng Yue "An Integrative Framework for Detecting Structural Variations in Cancer Genomes." *Nature Genetics*, 2018, In Press
- Yanli Wang, Bo Zhang, Lijun Zhang, **Lin An**, Jie Xu, Daofeng Li, Mayank NK Choudhary, Yun Li, Ming Hu, Ross Hardison, Ting Wang, Feng Yue. "The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions." *Genome Biology*, 2018, In Press
- Yu Zhang, **Lin An**, Feng Yue, Ross Hardison "Jointly characterizing epigenetic dynamics across multiple human cell types", *Nucleic Acids Research* 2016
- Meng, R., Zhang, J., **An, L.** et al. "Expression Profiling of Several Gene Families Involved in Anthocyanin Biosynthesis in Apple Skin During Fruit Development", *J Plant Growth Regul* 2016
- Boca SM, Nishida M, Harris M, Rao S, **Lin An**, Cheema AK, Gill K, et al. Discovery of Metabolic Biomarkers for Duchenne Muscular Dystrophy within a Natural History Study. *PLoS ONE* 2016
- D Wang, L Song, V Singh, S Rao, **Lin An**, S Madhavan "SNP2Structure: A public and Versatile Resource for Mapping and Three-Dimensional Modeling of Missense SNPs on Human Protein structures" *Computational and Structural Biotechnology Journal*, 2015
- Zhongtao Zhao, Huiquan Liu, Yongping Luo, Shanyue Zhou, **Lin An**, Chenfang Wang, Qiaojun Jin, Mingguo Zhou & Jin-Rong Xu., "Molecular evolution and functional divergence of tubulin superfamily in the fungal tree of life." *Scientific Reports*. 2014

In progress

- **Lin An**, Tao Yang, Qunhua Li, Yu Zhang. Hierarchical Domain Structure Reveals the Divergence of Activity among TADs and Boundaries. *Biorxiv*, 2018.

Honors & Awards

2013-2013 National Science Foundation Bootcamp scholarship
 2012-2013 Merited-based scholarship, Georgetown University
 2008-2009 Excellent Student Leader, NWSUAF

Conference Presentations**Oral Presentations**

05/2016 Nuclear Organization & Function "A Computational Framework to Predict Chromatin Interaction Using Genomic and Epigenomic Data" Cold Spring Harbor Laboratory
 07/2018 Intelligent Systems for Molecular Biology (ISMB) "Hierarchical Domain Structure Reveals the Divergence of Activity among TADs and Boundaries" Chicago

Professional Service**10/2013-10/2014 Volunteer Researcher****Rare Genomics Institute**

Investigated DNA mutations of a patient who developed rare symptoms. Co-worked with clinicians to interpret result and improve diagnosis.

05/2015-Present Reviewer

Bioinformatics, Plos Computational Biology and International Journal of Computational Biology and Drug Design (IJCBD)

2017-Present Organizer

Annual Bioinformatics & Genomics Retreat at The Pennsylvania State University

Teaching Experience

08/2015 – 12/2015 Teaching Assistant

Statistics Department, Pennsylvania State University – State College, PA

Advisor: Kwame Kankam Course: Stat 200